文章编号：1001-9014(XXXX)XX-0001-09

DOI:10.11972/j.issn.1001-9014.XXXX.XX.001

A multi-attention mechanism U-Net neural network for image correction of PbS quantum dot focal plane detectors

WANG Han-Ting[1,2,3], DI Yun-Xiang[2,4], QI Xing-Yu[4], SHA Ying-Zhe[4], WANG Ya-Hui[4], YE Ling-Feng[4], TANG Wei-Yi[4], BA Kun[4,5], WANG Xu-Dong[2], HUANG Zhang-Cheng[4], CHU Jun-Hao[2], SHEN Hong[2*], WANG Jian-Lu[1,2,3,4*]

(1. State Key Laboratory of Infrared Physics, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China.;
2. College of Physics and Optoelectronic Engineering, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China.;
3. University of Chinese Academy of Sciences, Beijing 100049, China.;
4. State Key Laboratory of Integrated Chips and Systems, Frontier Institute of Chip and System, Fudan University, Shanghai 200433, China.;
5. School of Microelectronics, Shanghai University, Shanghai 200444, China.)

Abstract: Near-infrared image sensors are widely used in fields such as material identification, machine vision, and autonomous driving. Lead sulfide colloidal quantum dot-based infrared photodiodes can be integrated with silicon-based readout circuits in a single step. Based on this, we propose a photodiode based on an n-i-p structure, which removes the buffer layer and further simplifies the manufacturing process of quantum dot image sensors, thus reducing manufacturing costs. Additionally, for the noise complexity in quantum dot image sensors when capturing images, traditional denoising and non-uniformity methods often do not achieve optimal denoising results. For the noise and stripe-type non-uniformity commonly encountered in infrared quantum dot detector images, a network architecture has been developed that incorporates multiple key modules. This network combines channel attention and spatial attention mechanisms, dynamically adjusting the importance of feature maps to enhance the ability to distinguish between noise and details. Meanwhile, the residual dense feature fusion module further improves the network´s ability to process complex image structures through hierarchical feature extraction and fusion. Furthermore, the pyramid pooling module effectively captures information at different scales, improving the network´s multi-scale feature representation ability. Through the collaborative effect of these modules, the network can better handle various mixed noise and image non-uniformity issues. Experimental results show that it outperforms the traditional U-Net network in denoising and image correction tasks.

Key words: PbS quantum dot focal plane detector, convolutional neural networks, image denoising, U-Net

用于PbS量子点焦平面探测器图像校正的多注意力机制U-Net神经网络

王瀚霆[1,2,3], 狄云翔[2,4], 齐星宇[4], 沙英哲[4], 王亚辉[4], 叶凌枫[4], 唐唯译[4], 巴 坤[4,5], 王旭东[2], 黄张成[4], 褚君浩[2], 沈 宏[2*], 王建禄[1,2,3,4*]

(1. 中国科学院上海技术物理研究所 红外物理国家重点实验室,上海 200083;
2. 中国科学院大学 杭州高等研究院物理与光电工程学院,杭州 310024;
3. 中国科学院大学,北京 100049;
4. 复旦大学 芯片与系统国家重点实验室 芯片与系统前沿技术研究院,上海 200043;

Received date: 2025-05-21, **revised date**: 2025-11-03 **收稿日期**:2025-05-21,**修回日期**:2025-11-03

Foundation items: Supported by Natural Science Foundation of China (62105100).

Biography: Wang Hanting (1999-), master degree candidate. Research area involves analog CMOS readout circuits and image denoising. E-mail: wanghanting22@mails.ucas.ac.cn. Yunxiang Di (1992-), Ph. D. candidate. Research area quantum dot infrared image sensors. E-mail: diyunxiang@fudan.edu.cn. Wang Hanting, Di Yunxiang contributed equally.

* **Corresponding author**: E-mail: jianluwang@fudan.edu.cn; hongshen@mail.sitp.ac.cn

5. 上海大学微电子学院,上海 200444)

摘要:近红外图像传感器广泛应用于材料识别、机器视觉和自动驾驶等领域。基于硫化铅胶体量子点的红外光电二极管可以通过单一步骤与基于硅的读出电路集成。基于此,我们提出了一种基于n-i-p结构的光电二极管,去除了缓冲层,进一步简化了量子点图像传感器的制造工艺,从而降低了制造成本。此外,对于量子点图像传感器在捕获图像时的噪声复杂性,传统的去噪和非均匀性校正方法往往无法达到最佳去噪效果。针对红外量子点探测器图像中常见的噪声和条纹型非均匀性,开发了一种包含多个关键模块的网络架构。该网络结合了通道注意力和空间注意力机制,动态调整特征图的重要性,以增强区分噪声和细节的能力。同时,残差密集特征融合模块通过分层特征提取和融合,进一步提高了网络处理复杂图像结构的能力。此外,金字塔池化模块有效地捕捉不同尺度的信息,提高了网络的多尺度特征表示能力。通过这些模块的协同作用,网络能够更好地处理各种混合噪声和图像非均匀性问题。实验结果表明,它在去噪和图像校正任务中优于传统的U-Net网络。

关 键 词:硫化铅量子点焦平面探测器;卷积神经网络;图像去噪;U-Net

Introduction

With the increasing applications in industrial inspection[1], material sorting[2], and autonomous drivin[3, 4], the field of infrared sensing and imaging has attracted widespread attention from researchers. Traditional InGaAs and HgCdTe photodetectors require high-quality single-crystal substrates for fabrication and rely on indium bump flip-chip bonding to integrate with complementary metal - oxide - semiconductor (CMOS) readout integrated circuits (ROIC). This not only increases the manufacturing cost and reduces the throughput but also significantly affects the resolution of focal plane photodetector arrays, thereby limiting their application scenarios[5]. Lead sulfide (PbS) colloidal quantum dot image sensors can be fabricated by directly spin-coating quantum dot materials onto the surface of the ROIC, eliminating the need for flip-chipbonding. Moreover, they can operate at room temperature, offering advantages in terms of manufacturing cost and array size scalability.

PbS quantum dot infrared image sensors have emerged as a key research focus. Liu et al[6]. reported a near-infrared image sensor based on PbS quantum dots with an array size of 640×512[6]. SWIR Vision Systems has developed multiple large-array PbS quantum dot cameras with resolutions of 1 920×1 080[7], covering several spectral bands from near-infrared to shortwave infrared, with a maximum detection range of up to 2 100 nm. However, these PbS quantum dot image sensors are typically based on p-i-n photodiodes[6, 8-10]. The implementation of the p-i-n structure generally requires the use of fullerene (C_{60}) as a buffer layer. While C_{60} materials exhibit excellent electron transport properties, they are expensive and involve complex fabrication processes[11], significantly increasing the overall structural complexity and manufacturing costs.

In addition, image processing methods for photodetectors have become a current research hotspot[12-15]. On one hand, the shot noise, 1/f noise, and other noises from the front-end photodetector will couple with the noise of the back-end readout circuit as well as the noise introduced during signal transmission. On the other hand, in different application scenarios, variations in temperature, light intensity, and signal transmission processes often result in the final captured images containing various forms and intensities of noise. Additionally, process deviations in CMOS readout circuits can lead to significant non-uniformity in the captured images, which is one of the main factors affecting the imaging quality of image sensors[16]. Currently, methods for direct image processing in both the spatial domain and the frequency domain have been widely studied. Traditional non-uniformity correction methods include calibration-based correction methods[17, 18]、Scene-based correction methods[19, 20] and filter-based correction methods [21, 22], etc. These methods can indeed achieve good results in image non-uniformity correction, but they also face some issues. Calibration-based methods assume system stability, but factors like temperature drift, vibrations, and changes in light or detector performance can cause sensor response drift, leading to calibration failures. These methods also require extensive data collection, which is time-consuming. Scene-based correction methods depend on the scene and may introduce errors with dynamic changes. Separating non-uniformity from scene information is challenging, especially in textured or high-frequency scenes, potentially distorting details. Filter-based methods smooth non-uniformity, but also blur image details, especially at edges, and assume uniform non-uniformity, which isn't always the case in real-world scenarios. After performing non-uniformity correction on an image, denoising is also a key step in improving the overall image quality. Traditional image denoising methods can generally be classified into spatial domain methods[23], frequency domain methods[24], and transform domain methods[25]. However, these methods may not be sufficient to capture all noise components when dealing with certain noise patterns. Additionally, the parameter selection for frequency domain filtering can be challenging, and these methods are highly sensitive to spectral characteristics. These factors, to varying degrees, limit the efficiency and application scenarios of photodetectors.

This paper proposes and develops a photodiode based on a normal-incidence n-i-p structure. Compared to the p-i-n structure, the n-i-p structure simplifies the manufacturing process and eliminates the need for the expensive C_{60} buffer layer. This not only reduces the overall manufacturing cost of the device but also simplifies the fabrication workflow, making it more suitable for large-

scale production and commercialization. We applied the detector to an array with a scale of 128×128, a pixel size of 15 μm, and a readout circuit chip fabricated using a 180 nm process, successfully fabricating a PbS quantum dot focal plane photodetector. Subsequently, we simulated potential noise scenarios that the quantum dot focal plane detector might encounter during actual use. Noise was artificially added to a public dataset, and based on UNet, we further designed an image processing algorithm that combines non-uniformity correction and image denoising for the quantum dot focal plane detector. This network model is a variant of UNet that incorporates various attention mechanisms and a residual dense feature fusion module, specifically designed for image denoising tasks in complex noisy environments. The goal is to improve the reconstruction quality and detail restoration capability of images under noise interference. Experiments demonstrate that this model achieves better denoising results in images with a mix of non-uniformity, image sensor noise, and quantum dot image sensor ring noise, significantly outperforming traditional UNet-based denoising methods.

1 Design and methods

To integrate with the readout circuit, quantum dot photodiodes require a normal-incidence structure. We designed an n-i-p normal-incidence structure, as shown in Figure 1(a), consisting of a multilayer configuration of Au/ZnO/PbS/PbS-EDT/ITO. In this structure, ZnO acts as the electron transport layer, PbS CQDs as the light-absorbing layer, PbS-EDT (The PbS CQD thin film, modified by 1, 2-ethanedithiol solid-phase ligand exchange,) as the hole transport layer, and Indium Tin oxide as the transparent conductive electrode. Compared to the p-i-n structure, the n-i-p structure simplifies the fabrication process and eliminates the need for the expensive C_{60} buffer layer. The transmission electron microscopy (TEM) image (Figure 1(b)) shows a cross-sectional view of our PbS photodiode. The thickness of each layer in the device is uniform and dense, with clear interfaces. The response wavelength range of this photodetector is 400-1 100 nm. To evaluate the photodiode's response to near-infrared light, we conducted current density – voltage (J-V) measurements under 940 nm laser illumination at different light intensities, as shown in Figure 1(c). Initially, under dark conditions without illumination, the photodiode's J-V curve exhibited clear diode rectification behavior. The rectification ratio was approximately five orders of magnitude, indicating excellent rectification performance. Furthermore, the device demonstrated a low dark current density, approximately 0.13 $\mu A \cdot cm^{-2}$ at a reverse bias of -0.1 V, which reflects a low leakage current and ensures a high signal-to-noise ratio. Under illumination, the device's photoresponse was tested across a range of 940 nm laser intensities, from 10 $\mu W \cdot cm^{-2}$ to 11 $W \cdot cm^{-2}$. The results demonstrate that the photodiode maintained a nearly constant response rate of approximately 0.3 A/W throughout the entire measurement range. Furthermore, under reverse bias, the photocurrent remained almost constant, indicating that the device maintains efficient photoelectric conversion even at low reverse voltages. This characteristic provides a strong foundation for the development of low-power, high-sensitivity PbS quantum dot image sensors.

The readout circuit chip structure, shown in Figure 2, includes a 128×128 pixel array, column-level amplifiers, output buffers, and row-column selection circuits. Each pixel measures 15 μm, with a full-well capacity of 6.366 Me^{-}.

Neural networks, through extensive training, learn noise distributions and image characteristics, enabling them to adapt to complex noise patterns and diverse scene properties without relying on fixed assumptions. This allows them to handle various types of noise and image non-uniformity, including mixed and time-dependent noise. Convolutional Neural Networks (CNNs) extract multi-level features, effectively capturing both local and global details, and preserving edges and textures better than traditional methods. By optimizing the input-output mapping through loss functions (such as MSSIM, MSE, L1), neural networks eliminate the need for intermediate processes to separate noise from signal, enabling global optimization. They can also update correction parameters in real-time in dynamic scenes, unaffected by time-de-

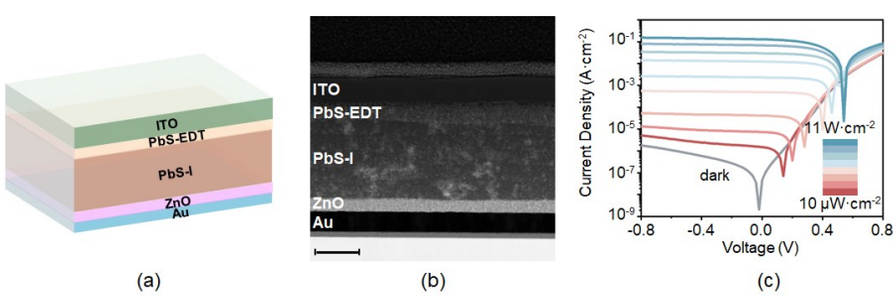

图 1 (a) PbS CQDs 光电二极管的示意图;(b) PbS CQDs 光电二极管的横截面 TEM 图像,比例尺为 100 nm;(c) 在暗环境和近红外(940 nm)照明下,设备的半对数 *J-V* 曲线,光强度范围从 0.2 $\mu W \cdot cm^{-2}$ 到 2 $W \cdot cm^{-2}$。

Fig. 1 (a) Schematic diagram of the PbS CQDs photodiode; (b) Cross-section TEM image of the PbS CQDs photodiode with a scale bar of 100 nm; (c) Semilog *J-V* curves of the device in the dark and under NIR (940 nm) illumination at intensities from 10 $\mu W \cdot cm^{-2}$ to 11 $W \cdot cm^{-2}$.

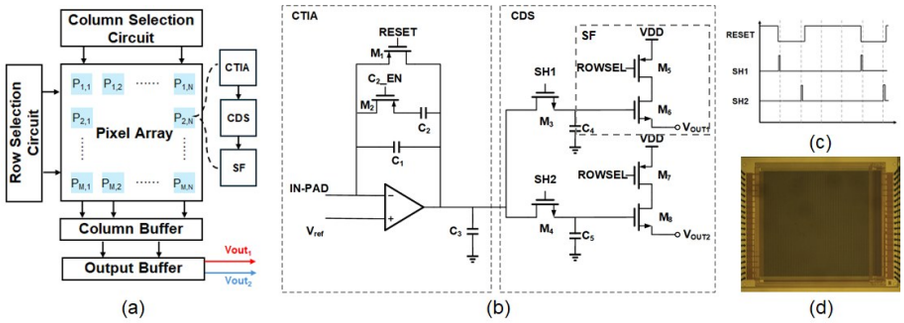

图2 （a）ROIC芯片的整体结构示意图，包括像素电路、行选择电路、列选择电路、列缓冲区和输出缓冲区；（b）像素电路结构示意图，包括CTIA、CDS和SF部分，具有两个输出信号V_{out1}和V_{out2}；（c）数字信号时序：SH1的上升沿和RESET的下降沿相重合，而SH2的下降沿应在RESET信号的上升沿来临之前；（d）芯片的实物图：该芯片基于180 nm工艺，阵列大小为128×128。芯片的顶部和底部分别有20个焊盘。

Fig. 2 （a）The overall structure diagram of ROIC chip, including the pixel circuit, row selection circuit, column selection circuit, column buffer, and output buffer;（b）Pixel circuit structure schematic, including CTIA, CDS, and SF sections, with two output signals, V_{out1} and V_{out2};（c）Digital signal timing: The rising edge of SH1 coincides with the falling edge of RESET, while the falling edge of SH2 should occur before the rising edge of the RESET signal;（d）Physical image of the chip: This chip is based on the 180 nm process with an array size of 128×128. There are 20 pads each on the top and bottom of the chip.

pendent drift or background changes, while leveraging GPU acceleration to enhance efficiency compared to traditional methods.

U-Net was first proposed in 2015 [26]. Currently, U-Net is widely used in various computer vision tasks [3], especially for problems that require pixel-level predictions, due to its efficient and accurate segmentation capabilities. U-Net adopts an Encoder-Decoder architecture, where the encoding process extracts features and downsamples the image through a series of convolutional layers and max-pooling layers, gradually capturing high-level semantic features of the input image. In the decoding process, upsampling and convolution layers are used to restore the feature maps and gradually recover the image′s spatial resolution, achieving fine segmentation. Some research has demonstrated that U-Net can also achieve good results in image denoising tasks [27, 28]. Javier et al. proposed RDUNet, a neural network that combines Residual Networks, Dense Networks, and the U-Net structure, designed for image denoising and other low-level computer vision tasks. Based on U-Net, we retained certain aspects of RDUNet′s residual dense feature fusion module and added the Pyramid Pooling module and Dual Attention module. The main network structure is shown in Figure 3.

We first introduced the Residual Dense Fusion Block, whose structure is shown in Figure 4(a). This is a network module that performs feature fusion using dense connections and residual mechanisms. The module enhances information flow by progressively extracting and merging features, while the residual connections preserve the original input features. The core of the module is a network made up of multiple Dense Connections. The input to each layer is the output of all previous layers (including the original input x). The input feature map at the i-th layer is a concatenation of the outputs of all previous i layers, followed by ReLU activation, which further enhances the information flow, allowing each layer to make full use of features from previous layers. The outputs from all layers are then concatenated into a larger

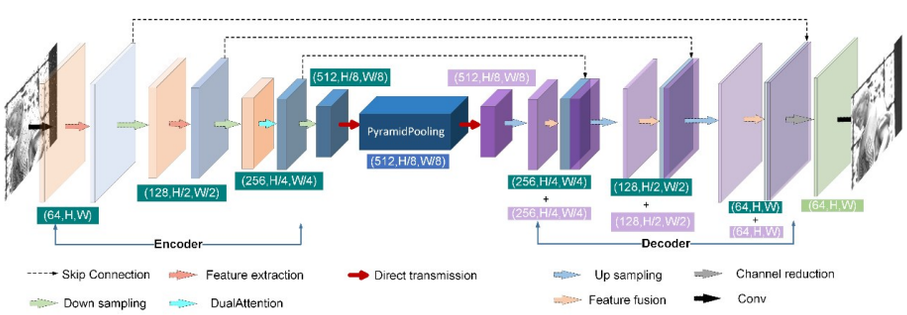

图3 DRPUNet网络的整体结构示意图。该网络在传统的UNet架构基础上，添加了金字塔池化模块，并在采样过程中融合了双重注意力机制和残差密集融合块

Fig. 3 The overall structure diagram of the DRPUNet network. This network builds upon the traditional UNet architecture by adding a Pyramid Pooling module and incorporates a Dual Attention mechanism and a Residual Dense Fusion Block during the sampling process

feature map, further enhancing the model's expressive power, allowing each layer to learn diverse information from different levels. The fused feature map is then processed by a 1×1 convolutional layer to produce the convolved output $x_{\text{fused\_final}}$. Finally, a residual connection adds the input x and $x_{\text{fused\_final}}$ together to obtain the final output x'. This process ensures that the original input x information is not lost in the network, helping to alleviate the vanishing gradient problem and making it easier for the network to learn effective features.

We introduced the Dual Attention module, whose structure is shown in Figure 4(b). This is a lightweight and efficient attention mechanism that focuses on input feature maps using two submodules: Channel Attention and Spatial Attention. It significantly improves network performance while maintaining low computational cost.

The Channel Attention module first applies average pooling along the channel dimension of the input image to generate channel weight coefficients. Then, two convolutional layers are used to compute the channel attention. Finally, the Sigmoid activation function is applied to obtain the normalized attention weights. X_c represents the feature map of the c-th channel of the input X, while W_1 and W_2 are the weights of the convolutional layers, and σ denotes the Sigmoid activation function.

$$\text{AvgPool}(X) = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{c,i,j} \quad , \quad (1)$$

$$\hat{z}_c = \sigma\left(W_2 \cdot \text{ReLU}\left(W_1 \cdot z_c\right)\right) \quad , \quad (2)$$

$$x_{\text{ca}} = x \cdot \hat{z}_c \quad , \quad (3)$$

The Spatial Attention module focuses on the importance of each position in the spatial dimension of the feature map, emphasizing the target regions and key pixels. It first applies average pooling and maximum pooling to each input channel to capture spatial information. The results of these two pooling operations are then concatenated together. A convolutional layer is used to compute the spatial attention weights. Finally, these spatial attention weights are applied to the output to produce the final weighted output x_{out}.

$$\text{avg\_out} = \frac{1}{C} \sum_{c=1}^{C} x_c \quad , \quad (4)$$

$$\text{max\_out} = \max_{c=1}^{C} x_c \quad , \quad (5)$$

$$x_{\text{out}} = x_{\text{ca}} \cdot \hat{a} \quad , \quad (6)$$

In the network's bottleneck, we introduced the Pyramid Pooling module (Figure 4(c)), which enhances the model's expressive capability by extracting multi-scale features through pooling at different scales. The module includes several steps: pooling at various scales, feature fusion, and convolutional output. First, we apply average pooling to the input feature map, followed by a convolution to reduce its channel number. These pooled feature maps are added to a feature list, forming a multi-scale set. Each pooled result is resized using Bilinear Interpolation to match the original input. The multi-scale features are then concatenated with the input and passed through a 1×1 convolution to obtain the final output feature map. The mathematical expression for this module is:

$$z_{\text{pool\_size}} = \text{AdaptiveAvgPool2d}\left(x, \text{pool\_size}\right) \quad , \quad (7)$$

$$z_{\text{conv}} = \text{Conv2d}\left(z_{\text{pool\_size}}, \frac{C}{|\text{pool\_sizes}|}, 1\right) \quad , \quad (8)$$

$$z_{\text{resized}} = \text{Interpolate}\left(z_{\text{conv}}, \text{size}(H,W), \text{mode} = \text{Bilinear}\right) \quad , \quad (9)$$

$$z_{\text{concat}} = \left[x, z_{\text{pool\_size}_1}, z_{\text{pool\_size}_2}, \cdots, z_{\text{pool\_size}_n}\right] \quad , \quad (10)$$

$$z_{\text{output}} = \text{Conv2d}\left(z_{\text{concat}}, C, 1\right) \quad , \quad (11)$$

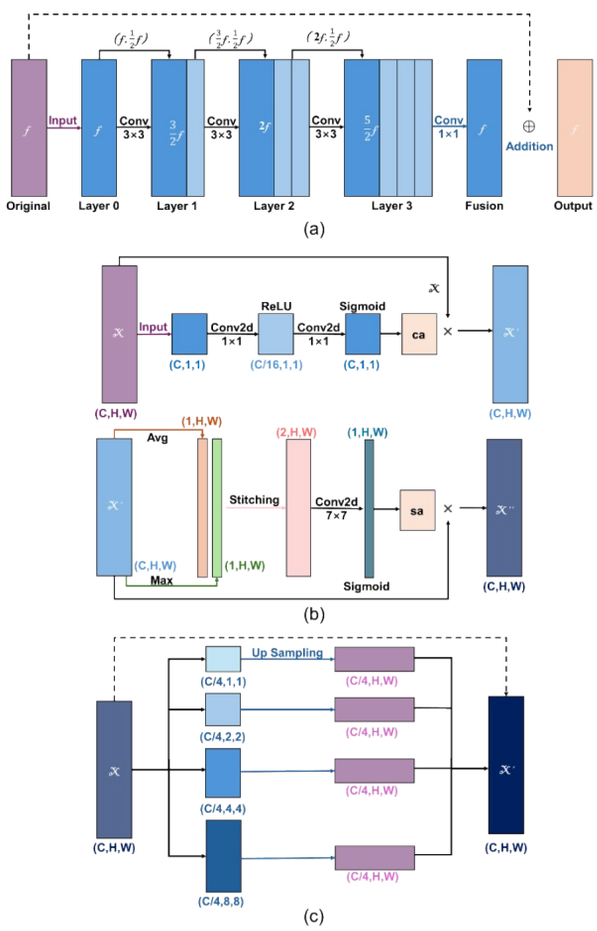

图4 (a)是残差密集融合块过程的示意图;(b)是双重注意力机制的示意图;(c)是金字塔池化模块的示意图

Fig. 4 (a) is the schematic diagram of the Residual Dense Fusion Block process; (b) is the schematic diagram of the Dual Attention mechanism, and (c) is the schematic diagram of the Pyramid Pooling module

The loss function is a method for quantifying the difference between the model's predictions and the true labels. During training, by minimizing the loss function, the model continuously adjusts its parameters to reduce prediction errors. For denoising tasks, the choice of loss function affects the model's sensitivity to noise and its ability to preserve details. MS_SSIM_L1_LOSS combines the advantages of MS-SSIM and L1 loss, enabling

structural perception and pixel difference constraints in the image. This ensures the reconstructed image is numerically closer to the target image, avoiding the color bias issue that may arise from using MS-SSIM alone. Therefore, we choose MS_SSIM_L1_LOSS as the loss function in this paper.

2 Experiments

We tested the model on an NVIDIA GeForce RTX™ 4 090 D GPU with an initial learning rate of 0.0001 and the Adam optimizer, which adapts the learning rate for each parameter, allowing faster convergence than SGD, especially in complex deep networks. Adam adjusts the learning rate dynamically, improving convergence speed and robustness. To train the model on diverse sensor noise, we added noise to the DIV2K dataset, simulating Gaussian, Poisson, and Salt-and-Pepper noise, combined with CMOS readout circuit non-uniformity. These types of noise were superimposed onto the images.

Gaussian noise follows a normal distribution, with values centered around a mean, and most values concentrated near it. It is primarily caused by low illumination, uneven brightness during image capture, and thermal or electronic noise from the photodetector. In images, this noise appears as random pixel value fluctuations, creating a grainy or blurry effect. The probability density function of Gaussian noise is defined by the variable z (pixel value), the mean μ (center of the noise curve), and the standard deviation σ, which determines the curve′s width.

$$p(z) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(z-\mu)^2}{2\sigma^2}} \quad , \quad (12)$$

Poisson noise is proportional to the signal intensity and primarily originates from the statistical fluctuation of photons. Specifically, the fundamental cause of Poisson noise is the randomness of photons hitting the pixel surface. At any given time, the number of photons hitting the pixel surface is a random variable that follows a Poisson distribution. This means that under the same lighting conditions, different pixels or the same pixel at different times will receive varying numbers of photons, leading to random changes in pixel values in the image. It is usually more noticeable in low light intensity scenes and causes different random noise levels between bright and dark areas of the image. The probability mass function formula is as follows, where K represents the number of events, and λ represents the average number of events occurring in a unit of time or space.

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad , \quad (13)$$

Salt-and-pepper noise appears as random black and white dots, where some pixels are extremally set to 0 (black) or 255 (white). This type of noise is usually associated with data transmission errors or image sensor malfunctions. In the image, it results in isolated bright or dark spots, with the surrounding pixels unaffected. The original pixel value $I(x,y)$ and the pixel value $I'(x,y)$ after salt-and-pepper noise can be represented by the following formula, where p_s and p_p are the probabilities that the pixel value becomes 0 and 255, respectively.

$$I'(x,y) = \begin{cases} 0 & \text{with probability } p_s \\ 255 & \text{with probability } p_p \\ I(x,y) & \text{with probability } 1 - p_s - p_p \end{cases} \quad , \quad (14)$$

In addition, the non-uniformity in the images captured by quantum dot focal plane detectors is typically caused by the column-level op-amp process deviations in the CMOS readout circuit. Since each column of pixels shares a single output buffer, manufacturing process deviations result in certain differences in the gain and offset of the column-level buffers[29]. This difference can be divided into DC and AC components. The DC component refers to fixed gain and offset differences across channels, while the AC component reflects channel drift over time, influenced by factors like temperature and environmental conditions. To construct the labels, non-uniformity can be simulated with randomly placed light and dark stripes of varying intensities. N represents the number of stripes (equal to the number of pixel columns), and k denotes the stripe number. This non-uniformity doesn't visually affect the image's brightness and rarely results in purely bright or dark stripes. Therefore, when simulating non-uniformity, it should be evenly distributed around (-255, 255). The following formula can be used for simulation:

$$\text{noisy}_{\text{image}(y,x)} = \max\left(-255, \min\left(255, \text{image}(y,x) + \sum_{k=0}^{N-1} v_k \cdot 1_{[x_k, x_k + W_s)}(x) \right) \right) \quad , \quad (15)$$

In quantum dot focal plane detector fabrication, spin-coating the quantum dot material onto the CMOS circuit creates ring-like noise patterns due to surface protrusions. These patterns resemble salt-and-pepper noise but have non-extreme pixel values and a nonlinear response to light. We classify them as pixel defect noise, which can reduce image quality by showing abnormal behavior under most lighting conditions.

When constructing the dataset, in addition to adding the above noise and stripe non-uniformity, we also added pixel defect noise to better match the images actually captured by the quantum dot focal plane detector. The construction process is shown in Figure 5. We randomly selected images from the DIV2K dataset, divided them into 128×128 sub-images, and then applied the noise addition method described above. Stripe non-uniformity (Figure 5(b)) and the aforementioned noise (Figure 5(c)) were added to the original image (Figure 5(a)). This resulted in a training set with 7,000 images and a validation set with 700 images, which were used for model training, validation, ablation experiments, and performance comparisons with other research works.

We performed denoising on the images actually captured by the PbS quantum dot photodetector, as shown in Figure 6. (a) is the raw image, where the circular and

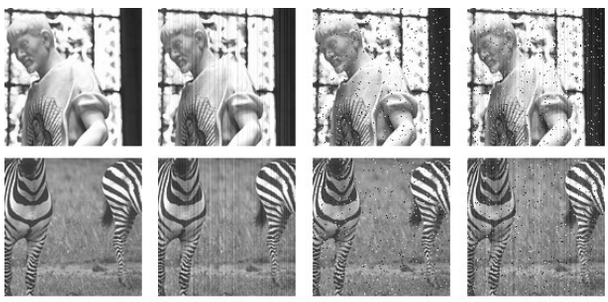

图5　这组图像展示了模型训练数据集的构建过程。我们从DIV2K数据集中选取了一些图像，转换为灰度图并裁剪为128×128大小。然后，我们模拟了PbS量子点探测器在实际应用中可能遇到的噪声类型，并将其人工添加到图像中，构建了训练集和验证集：(a)为原始图像；(b)为添加了非均匀噪声的图像；(c)为混合了高斯噪声、椒盐噪声、泊松噪声和均匀噪声的图像；(d)为将非均匀噪声与(c)中的混合噪声相结合的图像

Fig. 5　This set of images illustrates the construction of the training dataset for the model. We selected images from the DIV2K dataset, converted them to grayscale, and cropped them to 128×128 size. We then simulated the noise types the PbS quantum dot detector might encounter in real applications and added them to create the training and validation sets. (a) is the original image; (b) shows the image with added non-uniform noise; (c) features a mix of Gaussian, salt-and-pepper, Poisson, and uniform noise; (d) combines non-uniform noise with the mixed noise from (c)

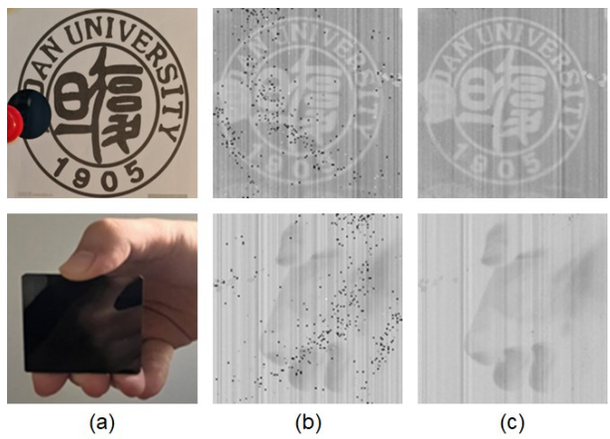

图6　这组图像展示了模型对实际由PbS焦平面探测器捕获图像的处理效果：(a) 是相机拍摄的真实照片，其中黑色圆形和方形代表硅片；(b) 是PbS焦平面探测器捕获的近红外图像；(c) 是模型最终处理后的图像

Fig. 6　This set of images demonstrates the model's processing effect on images actually captured by the PbS focal plane detector: (a) is a real photograph taken by the camera, with black circular and square shapes representing silicon wafers; (b) is the near-infrared image captured by the PbS focal plane detector; (c) is the final processed image by the model

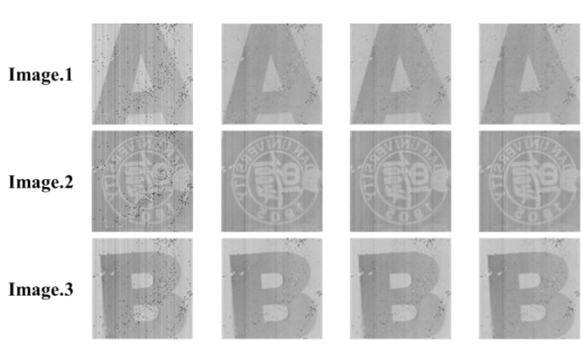

图7　使用了三张实际由PbS量子点光电探测器收集的图像(图像1、图像2、图像3)来测试RDUNet、UNet和BUNet模型

Fig. 7　Using three images actually collected by PbS quantum dot photodetectors (Image. 1, Image. 2, Image. 3) to test the RDUNet, UNet and BUNet models

square black masks are 940nm long-pass filters that block visible light and allow near-infrared light to pass through. (b) is the image captured by the PbS quantum dot photodetector under tungsten light illumination, where the tungsten light source contains infrared components. After filtering, the hand and the letters in the badge are clearly visible. However, the image still contains significant noise and stripe non-uniformity. We used the trained model to process (b), resulting in image (c). As seen, the noise in the actual captured image has been largely removed, the image non-uniformity has been significantly reduced, and the details and edges of the object being measured have been preserved as much as possible.

We also used the UNet and RDUNet networks to process the actual captured images, and the results are shown in Figure 7. For complex patterns like the badge, DRPUNet achieves better results, while for simpler patterns like the letters A and B, DRPUNet demonstrates better denoising performance for dense pixel defect noise compared to UNet and RDUNet.

We then used PSNR and SSIM as image evaluation metrics to assess the images processed by the proposed model. We used the BCDS public dataset as the test set. Since the array size of our detector′s readout circuit is 128×128, we selected a portion of images for cropping, resulting in a test set with 408 images and labels. Noise was artificially added to the test set in the same way as the training set to evaluate the model′s performance. Using the same hyperparameters, we denoised the dataset using UNet, RDUNet, and our proposed DRPUNet. The results are shown in Table 1. As seen, our model outperforms both UNet and RDUNet in denoising performance, while having fewer parameters than RDUNet, achieving better results in a shorter training time.

After confirming that our model outperforms others in denoising under the same training conditions, we conducted ablation experiments to verify the positive contri-

Table 1 Using the BCDS dataset to compare the denoising performance of UNet, RDUNet, and DRPUNet, with PSNR and SSIM as evaluation indicators for denoising effectiveness. After training for the same number of epochs with the same dataset, DRPUNet was able to achieve better denoising performance

表1 利用BCDS数据集对比了UNet、RDUNet和DRPUNet的去噪性能，以PSNR和SSIM作为去噪效果的评估指标。在使用相同数据集训练相同周期数后，DRPUNet能够实现更优的去噪性能

Method	PSNR	SSIM
UNet	33.067 8	0.966 0
RDUNet	36.630 7	0.982 8
DRPUNet	38.710 1	0.986 5

bution of each module. Since ablation studies on the traditional UNet Encoder-Decoder structure are widely available, we focused on separately adding modules to the UNet and combining them in pairs, comparing the results to the DRPUNet network (see Table 2). The results show that incorporating all three modules significantly improves network performance over UNet, with the best denoising performance achieved when all modules are included, outperforming RDUNet. This confirms the effectiveness and rationale behind these modules.

Table 2 Ablation experiment: The individual modules and their pairwise combinations are kept separately, trained under the same conditions, and tested with the same dataset. The results ultimately prove that the inclusion of these modules positively enhances the performance of the U-Net network, and the simultaneous introduction of all three modules leads to the best denoising performance of the model

表2 消融实验：分别保留各单独模块及其两两组合，在相同条件下进行训练，并使用相同数据集进行测试。结果最终证明，这些模块的引入对U-Net网络的性能有积极提升作用，同时引入所有三个模块能使模型的去噪性能达到最佳

Residual	Attention	ASPP	PSNR	SSIM
√			38.568 4	0.986 0
	√		33.451 8	0.958 2
		√	36.751 1	0.980 9
	√	√	35.312 5	0.978 7
√		√	38.540 7	0.986 2
√	√		38.671 9	0.986 3
√	√	√	38.710 1	0.986 5

3 Conclusions

This paper proposes a photodiode based on a normal-incidence n-i-p structure to simplify quantum dot image sensor manufacturing and reduce costs. It addresses both traditional noise and unique noise accumulation caused by the device fabrication process and circuit surface flatness. A simulated dataset of detector-captured images was created by modeling noise generation. A U-Net variant with multi-attention mechanisms, including contextual frequency-adaptive wavelet transforms, illumination-invariant frequency attention, and channel and spatial attention modules, is introduced. Additionally, a Noise-Selective Residual Learning Path (NSRL) with dynamic selection separates noise and image details. The MS_SSIM_L1_LOSS loss function ensures both structural perception and pixel difference constraints, offering high perceptual quality with low computational complexity.

While the proposed method performs well, it has some limitations. The goal of image denoising is to remove noise while preserving important details, requiring a balance between noise suppression and detail retention. Over-removal of noise can blur edges and lose details. Additionally, for practical use in quantum dot focal plane cameras, reducing computational parameters without sacrificing performance is essential. Developing a lightweight model for real-world deployment will be a key focus for future improvements.

References

[1] BAGAVATHIAPPAN S, LAKSHMI P V, JAYAKUMAR T, et al. Infrared thermography for condition monitoring – a review[J]. Infrared Physics & Technology, 2013, 60: 35-55.

[2] LIU T, DONG J, CHEN C, et al. RISIR: rapid infrared spectral imaging restoration model for industrial material detection in intelligent video systems [J]. IEEE Transactions on Industrial Informatics, 2024, 20(7): 9301-9312.

[3] CHOI J D, KIM M Y. A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles: its calibration and application[C]//2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN). Jeju: IEEE, 2021: 361-365.

[4] SONG H, LEE J, KIM S, et al. Short-wave infrared (SWIR) imaging for robust material classification: overcoming limitations of visible spectrum data[J]. Applied Sciences, 2024, 14(23): 10089.

[5] MALINOWSKI P E, BOMANS P, VAN DER SCHEER S, et al. Image sensors using thin-film absorbers[J]. Applied Optics, 2023, 62(17): F21-F30.

[6] LIU J, HE X, ZHANG Y, et al. A near-infrared colloidal quantum dot imager with monolithically integrated readout circuitry[J]. Nature Electronics, 2022, 5(7): 443-451.

[7] GREGORY C, WALKER M, LIU H, et al. Colloidal quantum dot photodetectors for large format NIR, SWIR, and eSWIR imaging arrays[J]. SID Symposium Digest of Technical Papers, 2021, 52(1): 982-986.

[8] ZHANG L, GAO H, WANG X, et al. High-performance and stable colloidal quantum dots imager via energy band engineering[J]. Nano Letters, 2023, 23(14): 6489-6496.

[9] ANDRESEN B F, JOHNSON W R, SCHAFFER S R, et al. Low-cost SWIR sensors: advancing the performance of ROIC-integrated colloidal quantum dot photodiode arrays [C]//Infrared Technology and Applications XL. Baltimore: SPIE, 2014, 9070: 90701I.

[10] ANDRESEN B F, JOHNSON W R, CARR R, et al. Room temperature SWIR sensing from colloidal quantum dot photodiode arrays[C]// Infrared Technology and Applications XXXIX. Baltimore: SPIE, 2013, 8704: 87041G.

[11] PAN Y, CHEN Y, LI D, et al. Advances in photocatalysis based on fullerene C60 and its derivatives: properties, mechanism, synthe-

sis, and applications [J]. Applied Catalysis B: Environmental, 2020, 265: 118580.
[12] ABBASS M Y, ZHANG Y, LIU J, et al. An efficient technique for non-uniformity correction of infrared video sequences with histogram matching[J]. Journal of Electrical Engineering & Technology, 2022, 17(5): 2971-2983.
[13] LI Y, LIU N, XU J. Infrared scene-based non-uniformity correction based on deep learning model[J]. Optik, 2021, 227: 166026.
[14] LI T, HE J, ZHANG M, et al. Non-uniformity correction of infrared images based on improved CNN with long-short connections [J]. IEEE Photonics Journal, 2021, 13(3): 7800113.
[15] CHEN X, WANG H, ZHAO L, et al. Infrared image denoising based on the variance-stabilizing transform and the dual-domain filter[J]. Digital Signal Processing, 2021, 113: 103039.
[16] PIPA D R, REIS F D, VIEIRA F H, et al. Recursive algorithms for bias and gain nonuniformity correction in infrared videos [J]. IEEE Transactions on Image Processing, 2012, 21(12): 4758-4769.
[17] SHENG M, XIE J, FU Z. Calibration-based NUC method in real-time based on IRFPA[J]. Physics Procedia, 2011, 22: 372-380.
[18] WANG H, ZHANG L, LIU Q, et al. An adaptive two-point non-uniformity correction algorithm based on shutter and its implementation [C]//2013 Fifth International Conference on Measuring Technology and Mechatronics Automation. Hong Kong: IEEE, 2013: 174-177.
[19] LV B, ZHAO H, XU W, et al. Statistical scene-based non-uniformity correction method with interframe registration [J]. Sensors, 2019, 19(24): 5504.
[20] HU B L, ZHANG Y, WANG J, et al. A novel scene-based non-uniformity correction method for SWIR push-broom hyperspectral sensors [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 131: 160-169.
[21] CAO Y, ZHANG J, HE Y, et al. A multi-scale non-uniformity correction method based on wavelet decomposition and guided filtering for uncooled long wave infrared camera[J]. Signal Processing: Image Communication, 2018, 60: 13-21.
[22] AVERBUCH A, LIRON G, BOBROVSKY B Z. Scene based non-uniformity correction in thermal images using Kalman filter [J]. Image and Vision Computing, 2007, 25(6): 833-851.
[23] PIYUSHBHAI P D, BHATTACHARYA S, PATEL N, et al. An analytical study of spatial domain image denoising techniques [J]. International Journal of Engineering Research and Technology, 2015, 4(5): 387-391.
[24] LI C. Research on image denoising method based on dual frequency domain transform[C]//2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). Chongqing: IEEE, 2024: 861-864.
[25] ROY V. Spatial and transform domain filtering method for image denoising: a review[J]. International Journal of Modern Education and Computer Science, 2013, 5(11): 41-49.
[26] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Munich: Springer, 2015: 234-241.
[27] GURROLA-RAMOS J, DALMAU O, ALARCON T E. A residual dense U-net neural network for image denoising [J]. IEEE Access, 2021, 9: 31742-31754.
[28] FAN C M, LIU T J, LIU K H. SUNet: swin transformer UNet for image denoising [C]//2022 IEEE International Symposium on Circuits and Systems (ISCAS). Austin: IEEE, 2022: 2333-2337.
[29] HE Z, LIU W, LIU C, et al. Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: a deep-learning approach[J]. Applied Optics, 2018, 57(18): D155-D164.